PATENT CLAIMS

We claim:

1. A computer implemented method of analyzing electronic data comprising the steps of:

a) providing a processing unit capable of receiving electronic data;

b) further providing a storage device coupled to said processing unit;

c) accessing one or more electronic data files, each said data file having a structure;

d) analyzing said one or more electronic data files to identify record break information contained therein;

e) utilizing said record break information, parsing said one or more data files into one or more electronic data records;

f) analyzing each of said electronic data records to identify field break information contained therein;

g) utilizing said field break information, parsing each of said data records into one or more data fields; and,

h) generating output data describing said structure of said one or more electronic data files.

2. The method of claim 1, further comprising the steps of:

repeating steps d) through g); and

utilizing said record break information and said field break information, updating said output data.

3. The method of claim 1 or 2, further comprising the step of:

storing said output data within said storage device.

4. The method of claim 1, further comprising the step of:

assigning a tokenized symbolic identifier to one or more of said data fields.

5. The method of claim 1, further comprising the step of:

providing a user interface through which a user may modify said output data, said user interface coupled to said storage device.

6. The method of claim 1, further comprising the step of:

utilizing said output data, generating a translation document capable of translating electronic documents into one or more predefined formats.

7. The method of claim 1, further comprising the steps of:

receiving modification instructions;

applying said modification instructions to one or more of said data fields; and

generating a first plurality of data files containing one or more modified data fields.

8. The method of claim 7, further comprising the step of:

testing said first plurality of data files.

9. The method of claim 1, further comprising the step of:

identifying a file type associated with each of said electronic data files.

10. The method of claim 1, further comprising the step of:

combining substantially similar electronic data files.

11. The method of claim 1, further comprising the steps of:

identifying one or more types of said electronic data records; and

analyzing said record type of each of said electronic data records to determine a degree of similarity.

12. The method of claim 11, further comprising the step of:

determining a cardinality for each said record type.

13. The method of claim 11, further comprising the step of:

determining a sequence of representation for each said record type.

14. The method of claim 11, further comprising the step of:

representing said degree of similarity of each said record type within said output data.

15. The method of claim 12, further comprising the step of:

representing said cardinality of each said record type within said data file.

16. The method of claim 13, further comprising the step of:

representing said sequence of representation for each said record type within said data file.

17. A computer readable medium comprising a plurality of instructions for analyzing computer intelligible electronic data which, when read by a computer system having a processing unit capable of receiving electronic data coupled to a storage device capable of storing electronic data, causes the computer to perform the steps of:

a) accessing one or more electronic data files, each said data file having a structure;

b) analyzing said one or more electronic data files to identify record break information contained therein;

c) utilizing said record break information, parsing said one or more data files into one or more electronic data records;

d) analyzing each of said electronic data records to identify field break information contained therein;

e) utilizing said field break information, parsing each of said data records into one or more data fields;

f) generating output data describing said structure of said one or more electronic data files; and,

18. The medium of claim 17, wherein said plurality of instructions causes the computer to perform the additional steps of:

repeating steps b) through e);

utilizing said record break information and said field break information, updating said output data.

19. The medium of claim 17 or 18, further comprising the step of:

storing said output data within said storage device.

20. The medium of claim 17, wherein said plurality of instructions causes the computer to perform the additional step of:

assigning a tokenized symbolic identifier to one or more of said data fields.

21. The medium of claim 17, wherein said plurality of instructions causes the computer to perform the additional step of:

providing a user interface through which a user may modify said output data, said user interface coupled to said storage device.

22. The medium of claim 17, wherein said plurality of instructions causes the computer to perform the additional step of:

utilizing said output data, generating a translation document capable of translating electronic documents into one or more predefined formats.

23. The medium of claim 17, wherein said plurality of instructions causes the computer to perform the additional steps of:

receiving modification instructions;

applying said modification instructions to one or more of said data fields; and

generating a first plurality of data files containing one or more modified data fields.

24. The medium of claim 23, wherein said plurality of instructions causes the computer to perform the additional step of:

testing said first plurality of data files.

25. The medium of claim 17, wherein said plurality of instructions causes the computer to perform the additional step of:

identifying a file type associated with each of said electronic data files.

26. The medium of claim 17, wherein said plurality of instructions causes the computer to perform the additional step of:

combining substantially similar electronic data files.

27. The medium of claim 17, further comprising the steps of:

identifying one or more types of said electronic data records; and

analyzing said record type of each of said electronic data records to determine a degree of similarity.

28. The medium of claim 27, further comprising the step of:

determining a cardinality for each said record type.

29. The method of claim 27, further comprising the step of:

determining a sequence of representation for each said record type.

30. The method of claim 27, further comprising the step of:

representing said degree of similarity of each said record type within said output data.

31. The method of claim 28, further comprising the step of:

representing said cardinality of each said record type within said data file.

32. The method of claim 29, further comprising the step of:

representing said sequence of representation for each said record type within said data file.

33. A computer system for analyzing computer intelligible electronic data comprising:

a processing unit for accessing one or more electronic data files, each said data file having a structure, for analyzing said one or more electronic data files to identify record break information contained within said files, for parsing said one or more data files into one or more electronic data records according to said record break information, for analyzing each of said electronic data records to identify field break information contained within said records, for parsing each of said data records into one or more data fields according to said field break information and for generating or output data describing said structure of said one or more electronic data files.

34. The computer system of claim 33, wherein said processing unit is further defined as being capable of updating said output data.

35. The computer system of claim 33, wherein said computer system further comprises a storage device, said processing unit being capable of storing said output data within said storage device.

36. The computer system of claim 33, wherein said processing unit is further defined as being capable of assigning a tokenized symbolic identifier to one or more of said electronic data fields.

37. The computer system of claim 33, wherein said computer system further comprises an interface through which a user may modify said output data, said interface being coupled to said processing unit.

38. The computer system of claim 33, wherein said processing unit is further defined as being capable of generating a translation document capable of translating electronic data into one or more predefined formats.

39. The computer system of claim 33, wherein said processing unit is further defined as being capable of receiving modification instructions, applying said modification instructions to one or more of said data fields and generating a first plurality of data files containing one or more modified data fields.

40. The computer system of claim 39, wherein said processing unit is further defined as being capable of testing said first plurality of data files.

41. The computer system of claim 33, wherein said processing unit is further defined as being capable of identifying a file type associated with each of said electronic data files.

42. The computer system of claim 33, wherein said processing unit is further defined as being capable of combining a first plurality of said electronic data files having a substantially similar structure.

43. The computer system of claim 33, wherein said record break information comprises one or more line termination characters.

44. The computer system of claim 33, wherein said record break information comprises one or more record break characters.

45. The computer system of claim 33, wherein said field break information comprises one or more character type transitions.

46. The computer system of claim 33, wherein said field break information comprises one or more character counts.

47. The computer system of claim 33, wherein said processing unit is further defined as being capable of identifying one or more types of said electronic data records and analyzing said types of said electronic data records to determine a degree of similarity.

48. The computer system of claim 47, wherein said processing unit is further defined as being capable of determining a cardinality for each said record type.

49. The computer system of claim 47, wherein said processing unit is further defined as being capable of determining a sequence of representation for each said record type.

50. The computer system of claim 47, wherein said processing unit is further defined as being capable of representing said degree of similarity of each said record type within said output data.

51. The computer system of claim 48, wherein said processing unit is further defined as being capable of representing said cardinality of each said record type within said data file.

52. The computer system of claim 49, wherein said processing unit is further defined as being capable of representing said sequence of representation for each said record type within said data file.